

Mark L. Shurtleff USB # 4666
UTAH ATTORNEY GENERAL
Raymond A. Hintze USB #1501
Chief Civil Deputy Attorney General
J. Mark Ward USB # 4436
Assistant Attorney General
236 State Capitol Building
Salt Lake City, Utah 84114
Telephone 801-538-1191
Facsimile 801-538-1121

Thomas R. Lee USB # 5991
1395 E. Canberra Dr.
Lindon, Utah 84042
Telephone 801-378-9024
Facsimile 801-378-5893

ORIGINAL
CLERK, U.S. DISTRICT COURT

29 AUG 01 PM 3:23

DISTRICT OF UTAH

BY: MM
DEPUTY CLERK

Gene C. Schaerr
Michael S. Lee USB #8108
Jay T. Jorgensen USB #8216
SIDLEY & AUSTIN
1722 Eye St., NW
Washington, DC 20006
Telephone 202-736-8000
Facsimile 202-736-8711

UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF UTAH, CENTRAL DIVISION

STATE OF UTAH, Michael O. Leavitt, Governor, :
Olene S. Walker, Lieutenant Governor, Mark. L. :
Shurtleff, Utah Attorney General, L. Alma Mansell, :
President of the Utah Senate, Martin R. Stephens, :
Speaker of the Utah House, Mike Dmitrich, Utah :
Senate Minority Leader, Ralph Becker, Utah House :
Minority Leader, Orrin G. Hatch, United States :
Senator, Robert F. Bennett, United States Senator, :
James V. Hansen, Member of Congress, :
Christopher B. Cannon, Member of Congress, :
James Matheson, Member of Congress, Blake J. :
Russom, Michael Wayne Andersen, and Brent & :
Jean McGhie, :
Plaintiffs, :
v. :
DONALD L. EVANS, Secretary of Commerce, and :
William G. Barron, Acting Director, United States :
Census Bureau, :
Defendants. :

DECLARATION OF LARA J.
WOLFSON, Ph.D., REGARDING
PLAINTIFFS' ACQUISITION OF
DATA ON IMPUTATION IN
CENSUS 2000.

Case No. 2:01-CV-292G

I, Lara J. Wolfson, do hereby declare the following:

1. I am currently employed as a Professor of Statistics at Brigham Young University in Provo, Utah.
2. I hold both a Ph.D. and an M.S. in Statistics from Carnegie Mellon University in Pittsburgh, Pennsylvania. I also hold a Bachelor's degree in Mathematics from Simon Fraser University in Burnaby, British Columbia.
3. I have authored over 30 scholarly articles in the field of statistics.
4. As a statistical consultant, I have also conducted numerous statistical studies and compiled statistical reports to accompany those studies.
5. A significant portion of my work – both as a scholar and as a consultant – has been devoted to evaluating the statistical accuracy and quality of national censuses.
6. An additional portion of my work, both as a scholar and as a consultant, has been in the area of statistical techniques for accounting for missing data; in particular, in the use of statistical imputation methods.
7. I am currently serving as the Past President of the Utah Chapter of the American Statistical Association.
8. I have been retained as a consultant in this case to advise Plaintiffs on the statistical implications of the U.S. Census Bureau's methodology in the 2000 decennial census.
9. The conclusions and opinions expressed in this Declaration are supported by data and statistical analysis that other experts in my field would consider reliable.
10. In preparing this Declaration, I have reviewed the following: (a) population data produced by the Census Bureau in connection with the 2000 census; (b) the process by which the Census Bureau apportions congressional representation after conducting each decennial census;

(c) numerous documents detailing the Census Bureau's methods in conducting Census 2000 as well as previous decennial censuses; (d) numerous scholarly articles on methods used in the decennial census; and (e) documents produced by Defendants in response to Plaintiffs' discovery requests.

11. Plaintiffs have asked me to respond to the Defendants' mischaracterization of the availability (and interpretability) of data released by the US Census Bureau on the use of imputation in the 2000 Census in the Federal Defendants Reply Memorandum (hereafter DOJ Reply) at 54-58, and North Carolina Intervenors' Reply Memorandum (NC Reply) at 22-24.

12. Federal defendants assert that the Plaintiffs did not exercise due diligence in researching whether or not imputation impacted Utah's apportionment in the House of Representatives. (*See* DOJ Reply at 55). This is not the case.

13. Until the Census Bureau released of the ACE reports and their supporting documentation after March 1, 2001, there was no public indication that (a) any imputations had been conducted to add people to the apportionment count, and (b) that if such imputations had been done, there was an appreciable number of them. In fact, as defendants' own declarant points out (Hogan Declaration at 34), both the number of imputations, and the percentage of the apportionment population attributable to persons added through imputation, had been steadily declining to 1990, to the point in the 1990 census, fewer than 54,000 persons were added to the apportionment count (less than .05% of the total apportionment population). *See* DOJ Discovery Response (Amended) at 17 (hereafter DOJ Disc. Resp.). This small number did not affect the apportionment in 1990; and there was every reason to believe, prior to March 1st, 2001, that if any imputations were included in the apportionment population, they were unlikely to affect apportionment.

14. This presupposition is supported by other Census Bureau documents, such as the report from the 1998 Census Dress Rehearsal, which indicates that the “standard” for unclassified housing units was to have less than 0.05% (the percentage in 1990) left unclassified:

“At the end of nonresponse followup, all units should have had a classification of occupied, vacant, or deleted. The percentage remaining unclassified reflected the success of nonresponse followup. *The standard for the percentage of cases which were unclassified -- that is, there was no housing unit status after nonresponse followup -- was less than 0.05 percent of the units on the census unedited file [10].* The standard for unclassified cases was consistent with similar data from the 1990 Census.” Report Card: Evaluations of the Standards for Success, Census 2000 Dress Rehearsal Report Card, February, 1999.” (emphasis in original).

15. Defendants seem to imply that the plaintiffs could have obtained data on the use of imputation in the Census through a FOIA request, discovery, or through calling the Census Bureau Number. *See* DOJ Reply at 56. They also imply that this data would have been given to plaintiffs had it been requested through these channels (“Somehow, it apparently never occurred to plaintiffs or to Dr. Wolfson to seek the information through the Census Bureau.” DOJ Reply at 58). Based on my experience with the Census Bureau, we would not have obtained this data any earlier than the April 9, 2001 date on which we did finally obtain, and verify, the data.

16. I arrive at this conclusion based on the length of time it took me to obtain data from a previous FOIA request. Defendants claim that my FOIA request of January 31, 2001 (attached hereto as Appendix A) was responded to. It was, but in a very dilatory fashion; the data did not arrive until March 26, 2001. I filed the FOIA request after having first attempted to obtain the data directly from the Census Bureau on January 30, 2001. (see Appendix B to this declaration, containing e-mails received from Violeta Vasquez of the Census Bureau).

17. In a phone conversation with Ms. Vasquez during this exchange, I was given to understand that any future attempts on my part to obtain anything other than a publicly available document (that I could identify by its specific title) would have to be made either through a discovery request or a FOIA request due to my involvement with the Utah v. Evans lawsuit.

18. Based on this, and prior experience with the Census Bureau, I concluded that unless I knew the exact name of a document or data file, that the turn-around time on any FOIA request would be at least 50 days, and that while I might succeed in obtaining documents through FOIA, I was unlikely to receive any data not in a specific published document. The standard practice of the Census Bureau, because of the time schedule to release data to the public, is to not distribute the data to individual requestors if the data is scheduled to be made public. Since I could not ascertain the name or title of the specific document or data file that contained the actual numbers of imputations, separately for each state, it seemed that other channels would be a more fruitful route to pursue.

19. During the first week of March, when the ESCAP (Executive Steering Committee for Accuracy and Coverage Evaluation Policy) reports were released, the first indication that was given that imputations were performed in numbers large enough to potentially affect apportionment was found in the Report of the Executive Steering Committee for Accuracy and Coverage Evaluation Policy. On p. 25, this document states "Census 2000 contains over five million records where imputation procedures were used to create all the information. These are referred to as whole person imputations." Because of the ambiguity of this description, it was impossible to tell from this document whether the use of imputation affected the apportionment of Representatives to Utah.

20. As defendants themselves acknowledge (DOJ Disc Rep at 3, *supra* 1), the Census Bureau has not always been consistent with the terms used to describe “count imputation.”¹ “Whole person imputations” could be count imputation, or the term could have referred to imputation for the purpose of describing characteristics such as age, race, or sex. In this particular instance, as we later discovered, it referred to the sum total of both characteristic and count imputations.

21. In the past, the Census Bureau has also been inconsistent in the use of the terms “imputation”, “whole-person imputation”, “substitution”, and “allocation” to refer to different types of imputation.

22. Because of these inconsistencies, the data in DSSD Report B-17 (AR C01010-28) was insufficient to make a determination of whether or not Utah had standing to make a claim that it was denied a seat in the congressional apportionment due to imputation. The first five tables in this report were based on data from the *preliminary housing unit unedited file* – and thus were not based on the final data file from which apportionment population totals were derived. Only the last table in the report (Table 6) was based on the final, edited file.

23. The columns in Table 6 of this report were entitled “Population in Housing Units”, “Percent Persons Substituted”, and “Percent Persons Unclassified”. The last two columns contained numbers given only to a precision of two decimal points. There were three difficulties in using these columns to try to ascertain if Utah had been injured by the use of count imputation in the 2000 Census.

24. First, the definitions of “percent persons substituted” and “percent persons unclassified” are ambiguous. It was possible that the total number of count imputations included components from both columns. Second, the precision of the percentages given was inadequate to calculate

¹ Following the nomenclature that seems to have been adopted in the course of the present litigation, I am using the term “count imputation” to refer to the imputation of the existence (and subsequently, all characteristics) of an

with accuracy the number of persons through each state added to the population count through count imputation, even had the correct definitions been known. Third, it was not clear how the percentages themselves had been calculated (in other words, did the “population in housing units” column include or exclude those for whom percentages were given in the other two columns?).

25. The lack of accuracy in this table is illustrated by considering the numbers given for 1990. In 1990, approximately 54,000 persons were added to the population count through count imputation. In the first line of Table 6 of DSSD B-17, the total population in housing units in 1990 is given as 242,012,129; and the percent persons unclassified is given as 0.02%. The product of those two numbers yields approximately 48402. The difference between this number and the “true” number could have been due to rounding error, or it could have been due to the definition of what constitutes a count imputation.

26. Thus, given the publicly available information on March 1, 2001, it was not discernible, from DSSD B-17 alone, that Utah had been harmed by the use of imputation in the Census.

27. Defendants further claim that “The Press Release dated March 13, 2001.... contained the actual count imputation numbers for both Utah and North Carolina” Defs. Reply Brief at 57, and further allege that the claim that the Census Monitoring Board did not have sufficiently detailed information on imputations is “belied by the March 13 Press Release”, Defs. Reply Brief at 58, *supra* 1. But the press release did not contain those numbers. It contained the numbers of “non-data defined persons” for Utah and North Carolina, which included both count and characteristic imputations. The total number of “non-data defined persons” referred to in that press release was 5.7 million – not the 1.2 million people that all parties in this lawsuit agree were included in the apportionment count solely through the use of count imputation.

28. I verified this in a phone conversation with Dr. Eugene Erickson of Temple University, who then gave me the data file from which he had produced the original calculations in the press release. While the data did not break down “non-data defined persons” into the necessary sub-categories, I examined these data to see if the allegations in the March 13 press release were correct. As defendants are surely aware, given the sensitivity of apportionment calculations to small shifts in population, it is insufficient to show merely that Utah’s priority number for a congressional seat would become higher than North Carolina’s; it is also necessary to show that no *other* state would also have a shift in their priority number for a seat that would then put them ahead of Utah. The information in the press release alone was insufficient to verify this.

29. Moreover, the category of “non-data defined” people is broad, and includes persons known to exist, and enumerated during the census, but for whom detailed characteristic data is missing. This could, for example, be the seventh member of a household, since standard enumeration forms only allow for characteristics to be given for the first six members of a household. The characteristics of individuals included in the Census are not relevant to the apportionment of congressional representation; so the category of “non-data defined persons” needed to be broken down into “characteristic” and “count” imputations. Dr. Erickson did not have, nor was he aware of the existence of, a data file containing such a breakdown, which he could access through his affiliation with the Census Monitoring Board, Presidential Members.

30. When plaintiff’s counsel did receive a data file on March 27, 2001, with the “non-data defined persons” broken down into sub-categories of “Data Defined”, “Imputations”, “Total Substitutions”, “Substitutions of 3,4,5”, and “Group Quarters”, it came with no documentation. I informed plaintiffs’ counsel that all I could conclude from examining this dataset was what the columns “Data Defined” and “Group Quarters” meant, but that the meaning of “Imputations”,

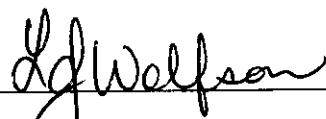
“Total Substitutions”, and “Substitutions of 3,4,5” was ambiguous at best, and all I could tell them was that it appeared that “Substitutions of 3,4,5” was a subset of “Total Substitutions” and concluded we needed to know exactly what those columns referred to in order to ascertain whether or not Utah had been injured by the use of imputation.

31. Because of the ambiguity in Census Bureau documentation in using a variety of overlapping definitions to the terms “imputations” and “substitutions” (a point conceded by the defendants in Defs. Disc. Repl. at 3, supra note 1), I could not definitively ascertain what the columns in the data file were (as it turns out, it was the column labeled “Substitutions of 3,4,5” that contained the numbers of persons added as “count imputations” into the Census).

32. Based on this, the only conclusion that can be reached is that plaintiffs’ in this lawsuit could not have obtained, at any earlier date than was done, an ascertainment of whether or not the use of count imputation in Census 2000 had cost Utah her fourth seat in Congress.

I hereby declare under penalty of perjury that the foregoing is true and correct.

Executed this 28 day of August, 2001.



Lara J. Wolfson, Ph.D.

Exhibits/
Attachments
to this document
have not been
scanned.

Please see the
case file.